The Astrophysical Journal, 769:63 (10pp), 2013 May 20 doi:10.1088/0004-637X/769/1/63

AN EFFICIENT APPROXIMATION TO THE LIKELIHOOD FOR GRAVITATIONAL WAVE STOCHASTIC BACKGROUND DETECTION USING PULSAR TIMING DATA

J. A. Ellis[1], X. Siemens[1], and R. van Haasteren[2]
[1] Center for Gravitation, Cosmology and Astrophysics, University of Wisconsin Milwaukee, Milwaukee, WI 53211, USA
[2] Max-Planck-Institut für Gravitationphysik (Albert-Einstein-Institut), D-30167 Hanover, Germany
Received 2013 February 8; accepted 2013 March 26; published 2013 May 6

ABSTRACT

Direct detection of gravitational waves by pulsar timing arrays will become feasible over the next few years. In the low frequency regime (10^{-7} Hz–10^{-9} Hz), we expect that a superposition of gravitational waves from many sources will manifest itself as an isotropic stochastic gravitational wave background. Currently, a number of techniques exist to detect such a signal; however, many detection methods are computationally challenging. Here we introduce an approximation to the full likelihood function for a pulsar timing array that results in computational savings proportional to the square of the number of pulsars in the array. Through a series of simulations we show that the approximate likelihood function reproduces results obtained from the full likelihood function. We further show, both analytically and through simulations, that, on average, this approximate likelihood function gives unbiased parameter estimates for *astrophysically realistic* stochastic background amplitudes.

Key words: gravitational waves – methods: data analysis – pulsars: general

1. INTRODUCTION

Gravitational waves (GWs) will very likely be detected in the next few years. Pulsar timing arrays (PTAs; Hobbs et al. 2010) as well as ground-based interferometers such as Advanced LIGO (Waldman 2011) are expected to make the first direct GW detection on a similar timescale, though they are sensitive to different and complementary regions of the GW spectrum. Ground-based instruments are most sensitive around 100 Hz, and the most promising sources at those frequencies are binaries of compact objects such as neutron stars and black holes (up to a few tens of solar masses). PTAs are most sensitive around 10^{-9} Hz, and the most promising source at those frequencies are supermassive binary black holes (SMBBHs) that coalesce when galaxies merge.

All the SMBBH mergers that have taken place throughout the history of our universe produce a stochastic background of GWs (Lommen & Backer 2001; Jaffe & Backer 2003; Wyithe & Loeb 2003; Volonteri et al. 2003; Enoki et al. 2004; Sesana et al. 2008; Sesana 2012; McWilliams et al. 2012), as well as individual periodic signals that may be detectable as above the confusion noise (Sesana et al. 2009; Sesana & Vecchio 2010; Roedig & Sesana 2012; Ravi et al. 2012; Mingarelli et al. 2012), and bursts (van Haasteren & Levin 2010; Cordes & Jenet 2012). A number of techniques have been implemented to search pulsar timing data for the stochastic background (Detweiler 1979; Stinebring et al. 1990; Lommen 2002; Jenet et al. 2005, 2006; Anholm et al. 2009; van Haasteren et al. 2009, 2011; Yardley et al. 2011; Cordes & Shannon 2012; Demorest et al. 2013), as well as periodic signals (Jenet et al. 2004; Yardley et al. 2010; Corbin & Cornish 2010; Lee et al. 2011; Ellis et al. 2012a, 2012b; Babak & Sesana 2012; Petiteau et al. 2013), and bursts (Finn & Lommen 2010).

For stochastic background searches, evaluations of the full likelihood are computationally challenging. PTAs are currently timing up to a few tens of pulsars, with several thousand points each. In addition, the likelihood function depends not only on the relatively small number of parameters that characterize GW stochastic background, but also on several intrinsic red and white noise parameters for each pulsar. A number of techniques have already been introduced to reduce the computational burden of such searches (van Haasteren 2013; Lentati et al. 2012; Taylor et al. 2013), and we will discuss these results later in the paper.

Although the stochastic background produces random changes in the times-of-arrival (TOAs) of an individual pulsar, the cross-correlation of its effects on two pulsars only[3] depends on the angular separation between pulsars (Hellings & Downs 1983). In this paper we introduce an efficient approximation to the likelihood by using an expansion to first order in the amplitude of the cross-correlation terms introduced by Anholm et al. (2009). This technique has already used to analyze the first International Pulsar Timing Array Mock Data Challenge (J. A. Ellis et al. 2013, in preparation). The approximation affords us a computational savings quadratic in the number of pulsars in the PTA, a factor of one to three orders of magnitude, depending on the size of the PTA.

This paper is organized as follows. In Section 2 we give an overview of the timing model. In Section 3 we write the likelihood function for the parameters of the stochastic background as well as intrinsic noise parameters of the pulsars, and introduce the first-order approximation in the amplitude of the cross-correlations. In Section 4, we show the effectiveness of our approximation using simulated gravitational wave backgrounds (GWBs), and that the level of bias introduced by our approximation is negligible for astrophysically reasonable stochastic background amplitudes. We conclude in Section 5 with a summary of our results, compare our results to other work to increase the computational efficiency of stochastic background searches (van Haasteren 2013; Lentati et al. 2012; Taylor et al. 2013), and introduce a technique that can be used to search for a combination of continuous wave signals and stochastic backgrounds, a possibility suggested by recent work (Ravi et al. 2012), which will be the basis for future work.

[3] In general, the effect of this cross correlation will depend on the individual pulsar terms. However, as was shown in Anholm et al. (2009), this contribution will become negligible for pulsar distances and GW frequencies of interest.

2. THE TIMING MODEL

In pulsar timing we measure the TOAs of radio pulses emitted from pulsars. These TOAs contain many terms of known functional form (pulsar period, spin-down, etc.), radiometer noise, pulse phase jitter, and possibly red noise either from interstellar medium effects, intrinsic pulsar spin noise (Shannon & Cordes 2010), or a stochastic GWB. Let the TOAs for a pulsar be given by

$$t^{\rm obs} = t^{\rm det}(\boldsymbol{\xi}_{\rm true}) + n, \tag{1}$$

where $t^{\rm obs}$ is the observed TOA, $t^{\rm det}$ is the deterministic modeled TOA parameterized by timing model parameters $\boldsymbol{\xi}_{\rm true}$, and n is the noise in the measurement which we will assume to be Gaussian. We will discuss the exact form of the covariance matrix for the noise n in the next section. Assuming we have an estimate of the true timing model parameters, $\boldsymbol{\xi}_{\rm est}$ (either from information gained when discovering the pulsar or past timing observations), then we can form the pre-fit residuals as follows

$$\begin{aligned}\delta t^{\rm pre} &= t^{\rm obs} - t^{\rm det}(\boldsymbol{\xi}_{\rm est}) = t^{\rm det}(\boldsymbol{\xi}_{\rm true}) - t^{\rm det}(\boldsymbol{\xi}_{\rm est}) + n \\ &= t^{\rm det}(\boldsymbol{\xi}_{\rm true}) - t^{\rm det}(\boldsymbol{\xi}_{\rm true} - \delta\boldsymbol{\xi}) + n \\ &\approx \left.\frac{\partial t^{\rm det}(\boldsymbol{\xi}_{\rm true})}{\partial \delta\boldsymbol{\xi}}\right|_{\delta\boldsymbol{\xi}=0} \delta\boldsymbol{\xi} + n + \mathcal{O}(\delta\boldsymbol{\xi}^2) \\ &\approx \left.\frac{\partial t^{\rm det}(\boldsymbol{\xi}_{\rm true})}{\partial \delta\boldsymbol{\xi}}\right|_{\delta\boldsymbol{\xi}=0} \delta\boldsymbol{\xi} + n \\ &= M\delta\boldsymbol{\xi} + n, \end{aligned} \tag{2}$$

where M is called the design matrix and we have assumed that our initial estimate of the model parameters is sufficiently close to the true values that we can approximate this as a linear system of equations in a small offset parameter $\delta\boldsymbol{\xi} \approx \boldsymbol{\xi}_{\rm true} - \boldsymbol{\xi}_{\rm est}$. In standard pulsar timing analysis, it is customary to obtain the best fit $\delta\boldsymbol{\xi}$ values through a weighted least squares minimization of the pre-fit residuals. In the most general case we should be performing a *generalized* least squares fit using a general covariance matrix for the noise n; however, in most cases we have no a priori knowledge of this covariance matrix and therefore assume that it is just diagonal with elements σ_i^2, where σ_i is the uncertainty of the ith TOA. Previous work (Coles et al. 2011) has used an iterative method to estimate the covariance matrix of the residuals and apply a generalized least squares fit, however; for this work we will only deal with residuals that have been created using a weighted least squares fit, since that is the standard procedure in pulsar timing residual generation. The value of χ^2 can be written in the following way (see Hobbs et al. 2006)

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{\delta t^{\rm pre}}{\sigma_i}\right)^2. \tag{3}$$

Defining $W = 1/\sigma_i$ we can minimize χ^2

$$\begin{aligned} 0 = \frac{\partial \chi^2}{\partial \delta\boldsymbol{\xi}} &= W^2 \left(M\delta\boldsymbol{\xi} + n\right) M^T \\ \Rightarrow M^T W^2 n &= -M^T W^2 M \delta\boldsymbol{\xi}, \end{aligned} \tag{4}$$

to obtain our best fit model parameters

$$\delta\boldsymbol{\xi}_{\rm best} = -(M^T W^2 M)^{-1} M^T W^2 n. \tag{5}$$

Here we have made the choice to include the weights, W, since TEMPO2 does a weighted fit and we want to reproduce the fitting procedure as accurately as possible. Finally we obtain the post fit residuals by substituting the best fit parameters into Equation (2)

$$\begin{aligned} r \equiv \delta t^{\rm post} &= M\delta\boldsymbol{\xi}_{\rm best} + n \\ \Rightarrow r &= Rn, \end{aligned} \tag{6}$$

where r is just shorthand notation for the post-fit residuals and

$$R = \mathbb{I} - M(M^T W^2 M)^{-1} M^T W^2, \tag{7}$$

is an oblique projection operator that transforms pre-fit to post-fit residuals and $\mathbb{I}$ is the identity matrix. All of the information about any noise source or stochastic GWB is encoded in n, however; we can never measure n directly because we must perform the timing model subtraction. Because of this we seek to work exclusively in terms of our observable quantities, r. It should be noted that in standard pulsar timing analysis this process must be iterated. In other words, we form pre-fit residuals from our initial guess of the parameters, we then minimize the χ^2 to get our best estimates of the parameters, however, this may not be a good fit because we have assumed that the pre-fit residuals are linear in the parameter offsets. Thus, we then form new parameter estimates from the best fit parameter offsets and iterate until the fit converges where the reduced χ^2 is used as our goodness of fit parameter.

3. THE LIKELIHOOD FUNCTION

The likelihood function for the timing residuals may be derived very simply from the likelihood of the underlying pre-fit Gaussian random processes. In this section we will derive an expression for the likelihood and introduce our approximation. We will also show that, in a frequentist sense, the maximum of the expectation value of the likelihood function is an unbiased estimator of the noise parameters in the low-signal regime.

Since we have assumed that our noise n is Gaussian and stationary, for a PTA with M pulsars we can write the probability distribution as the multi-variate Gaussian

$$p(\mathbf{n}|\boldsymbol{\theta}) = \frac{1}{\sqrt{\det(2\pi\,\boldsymbol{\Sigma}_n)}} \exp\left(-\frac{1}{2}\mathbf{n}^T \boldsymbol{\Sigma}_n^{-1}\mathbf{n}\right), \tag{8}$$

where

$$\mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_M \end{bmatrix} \tag{9}$$

is a vector of the noise time-series, $n_\alpha(t)$ for all pulsars, $\boldsymbol{\Sigma}_n$ is the *pre*-fit noise covariance matrix and $\boldsymbol{\theta}$ is a set of parameters that characterize the noise. Henceforth, a greek subscript will denote the pulsar number. As we noted above, we do not actually measure $\mathbf{n}$, we measure the timing residuals $\mathbf{r} = \mathbf{R}\mathbf{n}$ where

$$\mathbf{R} = \begin{bmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_M \end{bmatrix}. \tag{10}$$

We compute the likelihood for $\mathbf{r}$ as follows. Let

$$p(\mathbf{r}|\boldsymbol{\theta})d\mathbf{r} = p(\mathbf{n}|\boldsymbol{\theta})d\mathbf{n} \Rightarrow p(\mathbf{r}|\boldsymbol{\theta}) = p(\mathbf{n}|\boldsymbol{\theta}) \left|\frac{d\mathbf{n}}{d\mathbf{r}}\right|, \tag{11}$$

where $|\cdot|$ represents the determinant. We evaluate the Jacobian by *assuming* that $\mathbf{R}$ is invertible and writing $\mathbf{n} = \mathbf{R}^{-1}\mathbf{r}$, therefore

$$\left|\frac{d\mathbf{n}}{d\mathbf{r}}\right| = \left|\mathbf{R}^{-1}\right| = \frac{1}{|\mathbf{R}|} = \frac{1}{\sqrt{\det(\mathbf{R}\mathbf{R}^T)}}. \tag{12}$$

Substituting this result into Equation (11), we obtain

$$p(\mathbf{r}|\boldsymbol{\theta}) = \frac{1}{\sqrt{\det(2\pi\mathbf{R}\boldsymbol{\Sigma}_n\mathbf{R}^T)}} \exp\left(-\frac{1}{2}\mathbf{r}(\mathbf{R}^{-1})^T\boldsymbol{\Sigma}_n^{-1}\mathbf{R}^{-1}\mathbf{r}\right). \tag{13}$$

The product $\mathbf{R}\boldsymbol{\Sigma}_n\mathbf{R}^T$ is just the covariance matrix for the residuals

$$\boldsymbol{\Sigma} = \langle\mathbf{r}\mathbf{r}^T\rangle = \mathbf{R}\langle\mathbf{n}\mathbf{n}^T\rangle\mathbf{R}^T = \mathbf{R}\boldsymbol{\Sigma}_n\mathbf{R}^T, \tag{14}$$

so that the likelihood in terms of the timing residual data is simply

$$p(\mathbf{r}|\boldsymbol{\theta}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}\mathbf{r}^T\boldsymbol{\Sigma}^{-1}\mathbf{r}\right). \tag{15}$$

The inverse of $\boldsymbol{\Sigma}$ does not formally exist since we have removed degrees of freedom by fitting out the timing model. In practice, we can make use of a singular value decomposition (SVD) to compute the determinant and pseudoinverse to evaluate the likelihood. Viewed in this way, the likelihood function for the residuals is simply a change of coordinates where $\mathbf{R}$ is a linear (but not invertible) map from $\mathbf{n} \rightarrow \mathbf{r} = \mathbf{R}\mathbf{n}$.

The covariance matrix for the timing residuals is the block matrix,

$$\boldsymbol{\Sigma} = \begin{bmatrix} P_1 & S_{12} & \dots & S_{1M} \\ S_{21} & P_2 & \dots & S_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ S_{M1} & S_{M2} & \dots & P_M \end{bmatrix}, \tag{16}$$

where

$$P_\alpha = \left\langle r_\alpha r_\alpha^T \right\rangle, \tag{17}$$

$$S_{\alpha\beta} = \left\langle r_\alpha r_\beta^T \right\rangle\big|_{\alpha\neq\beta}, \tag{18}$$

are the auto-covariance and cross-covariance matrices, respectively, for each set of residuals. It is very important to note that we work *exclusively* in the post-fit variables. As above we use the post-fit residuals, $r_\alpha = R_\alpha n_\alpha$ and the post-fit auto- and cross-correlation matrices, $P_\alpha = R_\alpha P_\alpha^{\rm prefit} R_\alpha^T$ and $S_{\alpha\beta} = R_\alpha S_{\alpha\beta}^{\rm prefit} R_\beta^T$. Henceforth, we will drop any mention of pre-fit or post-fit as we will only work with post-fit variables.

It is worth pointing out that this treatment is somewhat different from previous Bayesian analyses (van Haasteren et al. 2009, 2011, the former hereafter VHML; van Haasteren & Levin 2010). We use a *conditional* probability distribution function (pdf) whereas VHML used a *marginalized* pdf. In other words, we fix the best fit parameter offsets, $\delta\boldsymbol{\xi}_{\rm best}$ through our use of the projection matrix R, whereas VHML marginalizes over the parameter offsets $\delta\boldsymbol{\xi}$ (see the Appendix for more details).

We would like to use the likelihood to determine the spectral index, $\gamma_{\rm gw}$, and amplitude, $A_{\rm gw}$, of the stochastic background from our data. The GW parameters are the same for all pulsars. In addition, each pulsar will have intrinsic noise parameters as well. The intrinsic pulsar timing noise is normally parameterized with four parameters: an amplitude A_α and spectral index γ_α for a power law red noise process, and EFAC and EQUAD parameters, $\mathcal{F}_\alpha$ and $\mathcal{Q}_\alpha$, for white noise processes. In general the EFAC parameter is a multiplicative factor representing any systematic effects in the uncertainty in each TOA based on the cross correlation of the folded pulse profile with a template (Taylor et al. 1992). The EQUAD parameter is an extra white noise parameter that is added to the TOA error in quadrature and could represent the expected pulse phase jitter (Cordes & Shannon 2010) and other white noise processes that are unaccounted for. Therefore, we write our auto-covariance as a sum of a common GWB term and a pulsar-dependent term

$$P_\alpha = N_\alpha + S_{a\alpha}, \tag{19}$$

where N_α is the intrinsic noise auto-covariance matrix and $S_{a\alpha}$ is the common GWB auto-covariance matrix for pulsar α. It is convenient to work in a block matrix notation where

$$\boldsymbol{\Sigma} = \mathbf{N} + \mathbf{S}_a + \mathbf{S}_c = \mathbf{P} + \mathbf{S}_c, \tag{20}$$

where $\mathbf{P}$ is a block diagonal matrix with diagonals P_α and $\mathbf{S}_c$ is block matrix with off diagonals $S_{\alpha\beta}$, and zero block matrices on the diagonal.

We will now quickly show that, in a frequentist sense, the maximum of the expectation value of the likelihood function is an unbiased estimator of our signal parameters $\boldsymbol{\theta} = \{A_{\rm gw}, \gamma_{\rm gw}, A_\alpha, \gamma_\alpha, \mathcal{F}_\alpha, \mathcal{Q}_\alpha\}$. We write the log likelihood function as

$$\ln\,\mathcal{L} = -\frac{1}{2}\left[\mathrm{Tr}\,\ln\boldsymbol{\Sigma} + \mathbf{r}^T\boldsymbol{\Sigma}^{-1}\mathbf{r}\right], \tag{21}$$

where we have used the fact that $\ln\det(A) = \mathrm{Tr}\ln(A)$ for a general matrix, A. To show that the maximum of the expectation value of this likelihood function is an unbiased estimator of the signal parameters, $\boldsymbol{\theta}$, we wish to show that it is maximized, on average, for signal parameters $\boldsymbol{\theta} = \boldsymbol{\theta}_{\rm true}$. Taking the expectation value, we obtain

$$\langle\ln\,\mathcal{L}\rangle = -\frac{1}{2}\mathrm{Tr}\left[\ln\boldsymbol{\Sigma} + \mathbf{X}\boldsymbol{\Sigma}^{-1}\right], \tag{22}$$

where $\mathbf{X} = \langle\mathbf{r}\mathbf{r}^T\rangle$ is the covariance matrix of the data. Defining $\partial_i = \partial/\partial\theta_i$, we obtain

$$\partial_i\langle\ln\,\mathcal{L}\rangle = -\frac{1}{2}\mathrm{Tr}\left[\boldsymbol{\Sigma}^{-1}\partial_i\boldsymbol{\Sigma} - \mathbf{X}\boldsymbol{\Sigma}^{-1}\partial_i\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\right]. \tag{23}$$

Assuming that our noise model is correct, we have $\mathbf{X} = \boldsymbol{\Sigma}$ and

$$\partial_i\langle\ln\,\mathcal{L}\rangle = -\frac{1}{2}\mathrm{Tr}\left[\boldsymbol{\Sigma}^{-1}\partial_i\boldsymbol{\Sigma} - \partial_i\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\right] = 0, \tag{24}$$

where we have used the fact that $\mathrm{Tr}(AB) = Tr(BA)$ for general matrices, A and B. Therefore, the maximum of the expectation value of the likelihood function is an unbiased estimator of our model parameters $\boldsymbol{\theta}$.

3.1. Likelihood with First-order Approximation

In practice the matrix $\boldsymbol{\Sigma}$ is quite large and therefore computationally prohibitive to invert. Since many multi-frequency residual data sets now have on the order of 10^3 points, for many modern PTAs the matrix $\boldsymbol{\Sigma}$ will be of the order of $10^4 \times 10^4$. We would like to avoid inverting the full covariance matrix if at all possible. First let us rewrite the cross-covariance as $\mathbf{S}_{c,\alpha\beta} = \zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}$, where $\mathbf{S}_{\alpha\beta}$ is the temporal cross covariance

between pulsar α and pulsar β. The coefficients represent the spatial correlations and are given by the Hellings and Downs coefficients

$$\zeta_{\alpha\beta} = \frac{3}{2}\frac{1-\cos\xi_{\alpha\beta}}{2}\ln\left(\frac{1-\cos\xi_{\alpha\beta}}{2}\right) - \frac{1}{4}\frac{1-\cos\xi_{\alpha\beta}}{2} + \frac{1}{2} + \frac{1}{2}\delta_{\alpha\beta}, \quad (25)$$

where $\xi_{\alpha\beta}$ is the angular separation of pulsars α and β, and $\delta_{\alpha\beta}$ is the Kronecker delta. We denote $\mathbf{P} = \delta_{\alpha\beta}\mathbf{P}_{\alpha\beta}$ as the auto-covariance matrix of pulsar α describing the noise and auto-covariance of the GWB. We then use the following notation to form matrices from indexed quantities: $\mathbf{P} = \{P_{\alpha\beta}\}$. Now, we perform the expansion of $\mathbf{\Sigma}^{-1}$ in terms of the coefficients $\zeta_{\alpha\beta}$

$$\begin{aligned}\mathbf{\Sigma}^{-1} &= (\mathbf{P} + \{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\})^{-1} = (\mathbb{I} + \mathbf{P}^{-1}\{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\})^{-1}\mathbf{P}^{-1} \\ &\approx \mathbf{P} - \left\{\sum_{\beta,\mu}\zeta_{\beta\mu}\mathbf{P}^{-1}_{\alpha\beta}\mathbf{S}_{\beta\mu}\mathbf{P}^{-1}_{\mu\nu}\right\} \\ &\quad + \left\{\sum_{\beta,\mu,\nu}\zeta_{\beta\mu}\zeta_{\mu\nu}\mathbf{P}^{-1}_{\alpha\beta}\mathbf{S}_{\beta\mu}\mathbf{P}^{-1}_{\mu\mu}\mathbf{S}_{\mu\nu}\mathbf{P}^{-1}_{\nu\sigma}\right\} + \mathcal{O}(\zeta^3).\end{aligned} \quad (26)$$

It is also possible to expand the determinant term in a similar fashion

$$\begin{aligned}\ln\det\mathbf{\Sigma} &= \mathrm{Tr}\ln\mathbf{\Sigma} = \mathrm{Tr}\ln(\mathbf{P} + \{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\}) \\ &= \mathrm{Tr}\left[\ln\mathbf{P} + \ln(\mathbb{I} + \mathbf{P}^{-1}\{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\})\right] \\ &\approx \mathrm{Tr}\left[\ln\mathbf{P} + \mathbf{P}^{-1}\{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\}\right. \\ &\quad \left. - \left\{\sum_{\beta,\mu,\nu}\zeta_{\beta\mu}\zeta_{\mu\nu}\mathbf{P}^{-1}_{\alpha\beta}\mathbf{S}_{\beta\mu}\mathbf{P}^{-1}_{\mu\mu}\mathbf{S}_{\mu\nu}\mathbf{P}^{-1}_{\nu\sigma}\right\}\right] + \mathcal{O}(\zeta^3).\end{aligned} \quad (27)$$

Here, the order $\mathcal{O}(\zeta)$ term is zero because $\mathbf{P}$ is block diagonal and $\{\mathbf{S}_{\alpha\beta}\}$ is block traceless and the trace of the product of a diagonal matrix and traceless matrix vanishes. If we ignore all terms of ζ^2 and higher order and return to our original notation, then we see that

$$\mathbf{\Sigma}^{-1} \approx \mathbf{P}^{-1} - \mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1} + \mathcal{O}(\zeta^2) \quad (28)$$

$$\ln\det\mathbf{\Sigma} \approx \mathrm{Tr}\ln\mathbf{P} + \mathcal{O}(\zeta^2). \quad (29)$$

This derivation may give us the sense that this expansion may hold true for all GWB amplitudes; however, this is not true as we will now show. Although we have written this approximation in terms of an expansion in the Hellings and Downs coefficients, it is also useful to think of it as an expansion in the amplitude of the GWB. Indeed, that it how it was conceived of in Anholm et al. (2009). We have not performed a true first-order expansion, however, since the inverse of the auto-correlations matrix $\mathbf{P}^{-1} = (\mathbf{N} + A^2_{\rm gw}\mathbf{A}_a)^{-1}$ contains terms of infinite order in the amplitude. We can essentially think of the $\mathcal{O}(\zeta)$ terms in Equations (28) and (29) as the corrections to the amplitude parameter when we have a spatially correlated signal. Thus, we have truncated these correction terms at $\mathcal{O}(A^2_{\rm gw})$ and we would not expect this approximation to hold as $A_{\rm gw}$ becomes large with respect to the intrinsic noise in the pulsar as we will show in Section 4. With these approximations, it is now possible to write the approximate log-likelihood

$$\begin{aligned}\ln\mathcal{L} &= -\frac{1}{2}[\mathrm{Tr}\ln\mathbf{P} + \mathbf{r}^T\mathbf{P}^{-1}\mathbf{r} - \mathbf{r}^T\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1}\mathbf{r}] \\ &= -\frac{1}{2}\sum_{\alpha=1}^{M}\left[\mathrm{Tr}\ln P_\alpha + r_\alpha^T P_\alpha^{-1} r_\alpha \right. \\ &\quad \left. - \sum_{\beta\neq\alpha}^{M} r_\alpha^T P_\alpha^{-1} S_{\alpha\beta} P_\beta^{-1} r_\beta\right].\end{aligned} \quad (30)$$

In the second line we have explicitly written out the sum over pulsars and pulsar pairs in order to highlight the fact that we only need to invert the individual auto-covariance matrices as opposed to the inverse of the full block covariance matrix, thereby, significantly reducing the computational cost of a single likelihood evaluation. Consider a PTA with M pulsars with N TOAs each. For a full likelihood evaluation we must perform one Cholesky inversion of the full covariance matrix which scales like $\sim\alpha(MN)^3$ and $\sim M^2$ matrix multiplications which scale like $\sim\beta N^3$. However, one evaluation of the first-order likelihood requires M Cholesky inversions which scale like $\sim\alpha N^3$ and M matrix multiplications which, again, scale like $\sim\beta N^3$. Through benchmarking tests we have found that $\beta \sim 10\alpha$ and thus the matrix multiplications will dominate both likelihood calls for reasonable sized PTAs ($M \lesssim 100$) resulting in a computation speedup factor of $\sim(\alpha/\beta)M^2$.

It is possible to analytically show that the maximum of the expectation value of this approximate likelihood is an unbiased estimator in the same manner as above. First we take the expectation value of the log-likelihood

$$\langle\ln\mathcal{L}\rangle = -\frac{1}{2}\mathrm{Tr}\left[\ln\mathbf{P} + \mathbf{X}\mathbf{P}^{-1} - \mathbf{X}\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1}\right] \quad (31)$$

and then take a derivative with respect to a model parameter

$$\begin{aligned}\partial_i\langle\ln\mathcal{L}\rangle = -\frac{1}{2}\mathrm{Tr}\left[\mathbf{P}^{-1}\partial_i\mathbf{P} - \mathbf{X}\mathbf{P}^{-1}\partial_i\mathbf{P}\mathbf{P}^{-1}\right. \\ + \mathbf{X}\mathbf{P}^{-1}\partial_i\mathbf{P}\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1} - \mathbf{X}\mathbf{P}^{-1}\partial_i\mathbf{S}_c\mathbf{P}^{-1} \\ \left. + \mathbf{X}\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1}\partial_i\mathbf{P}\mathbf{P}^{-1}\right].\end{aligned} \quad (32)$$

Here we will work in the small signal regime where $A^2_{\rm gw}$ is small compared to the amplitude of the intrinsic noise. Assuming that we have modeled the covariance matrix correctly, we have $\mathbf{X} = \mathbf{\Sigma}$. Writing out the explicit amplitude dependence, we assume

$$\mathbf{P} = \mathbf{N} + A^2_{\rm gw}\mathbf{A} \Rightarrow \mathbf{P}^{-1} \approx \mathbf{N}^{-1} - A^2_{\rm gw}\mathbf{N}^{-1}\mathbf{A}\mathbf{N}^{-1} \quad (33)$$

$$\mathbf{\Sigma} = \mathbf{N} + A^2_{\rm gw}\mathbf{A} + A^2_{\rm gw}\mathbf{C}, \quad (34)$$

where $\mathbf{N}$, $\mathbf{A}$, and $\mathbf{C}$ are the auto-covariance of the noise, the auto-covariance of the GWB, and the cross-covariance of the GWB, respectively. Then, to first order in $A^2_{\rm gw}$ we have

$$\begin{aligned}\partial_i\langle\ln\mathcal{L}\rangle = -\frac{1}{2}\mathrm{Tr}\left[\mathbf{N}^{-1}\partial_i\left(A^2_{\rm gw}\mathbf{A}\right)\right. \\ \left. - \mathbf{N}^{-1}\partial_i\left(A^2_{\rm gw}\mathbf{A}\right) - \partial_i\left(A^2_{\rm gw}\mathbf{C}\right)\mathbf{N}^{-1}\right] = 0,\end{aligned} \quad (35)$$

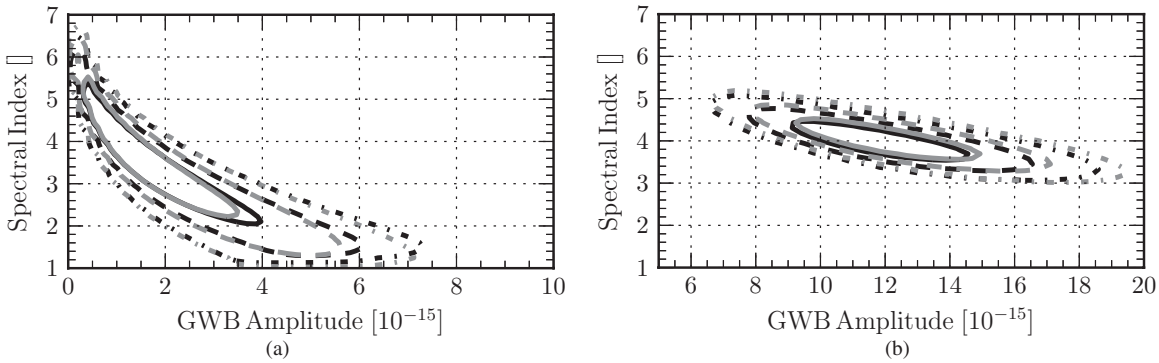

Figure 1. Comparison of full likelihood (gray) of VHML and the first-order likelihood (black). (a): 10 pulsars $A = 1 \times 10^{-15}$, (b): 10 pulsars $A = 1 \times 10^{-14}$.

where the first two terms cancel and the third term is the trace of the product of a diagonal matrix and a traceless matrix. Thus, to first order in $A^2_{\rm gw}$, the maximum of the expectation value of this approximate likelihood is an unbiased estimator of the our signal parameters θ in the weak signal limit. In other words, we have shown analytically that for reasonably small stochastic background amplitudes, as is expected, the parameters that we infer from this likelihood function will be unbiased on average. This result will be verified in the next section through the use of simulations.

4. SIMULATIONS

Here we will compare our first-order likelihood approximation to the full likelihood of VHML and perform mock searches of simulated data with and without an injected stochastic GWB in order to demonstrate its efficacy. We will also perform Monte Carlo simulations to test the consistency of our likelihood function. These simulations are solely meant as a proof of principle and do not claim to reproduce all features of real PTA data (irregular sampling, jumps, time varying Dispersion Measure (DM) corrections, etc.). However, our analysis method makes no assumptions about sampling by operating in the time domain and takes all timing model parameters into account via the projection matrices introduced in Section 2. The application of this method to real NANOGrav and IPTA data sets will be the subject of future work. For all simulations in the present work we use TEMPO2 and the `fake`, `GWbkgrd`, `general2` and `designmatrix` plugins to generate the residuals and the corresponding design matrices. All simulated white noise is solely radiometer noise at the level of 100 ns unless otherwise noted.

4.1. Mock Searches

First we will perform a simple test to compare the first-order likelihood of this work and the full likelihood of VHML. Here we use a PTA with 10 pulsars observed at a cadence of 20 TOAs per year for 5 years where we have fixed the EFAC parameter to be one (all white noise is encompassed in error bars as simulated) and assume that there is no intrinsic red noise, resulting in a search over two parameters; the amplitude of the stochastic GWB, A, and the power spectral index, γ. For both cases a grid search was carried out with 100 points in each dimension and $A \in (0, 1 \times 10^{-14})$ for an injected value of $A = 1 \times 10^{-15}$ and $A \in (0, 2 \times 10^{-14})$ for an injected value of $A = 1 \times 10^{-14}$, all the while we have $\gamma \in [1, 7]$. The results are presented in Figure 1 where the contours denote the one, two, and three sigma credible regions, the gray contours are from the VHML likelihood function, and the black contours are from the first-order likelihood.

In Figure 1(a) we have injected a stochastic GWB with $A = 1 \times 10^{-15}$ and $\gamma = 13/3$. First we notice that the injected value is well within the 1σ credible regions for both likelihood functions. We also see that the confidence contours are nearly identical, with the first-order likelihood preferring slightly larger amplitudes and smaller spectral indices. This simulation indicates that the first-order likelihood is a very good approximation to the full likelihood when our signal is relatively small, showing no discernible bias and faithfully reproducing nearly identical credible regions.

In Figure 1(b) we have injected a stochastic GWB with $A = 1 \times 10^{-14}$ and $\gamma = 13/3$. Again, the injected value lies within the 1σ credible region, however; now we do note a difference between two credible regions from the full and first-order likelihoods. The first-order likelihood is biased toward lower amplitudes and lower spectral indices. In fact, we can almost see where the first-order approximation begins to break down. Note that the contours are nearly identical for lower amplitudes and deviate more with increasing amplitude. This behavior is not surprising in that we know that this likelihood is only unbiased to first order in the amplitude as shown in Section 3. In fact, it is impressive that this approximation performs this well with only a small bias in the large signal limit (even with timing residuals lower than 100 ns in many pulsars, the signal-to-noise-level of the data simulated here is well above any reasonable estimates for future PTA sensitivities). This bias will be discussed further in Section 4.3.

The simulations used in the work have been quite ideal and do not contain any systematic effects such as clock errors which can manifest as a correlated noise sources with uniform correlation coefficients (Yardley et al. 2011), errors in solar system ephemerides, which can manifest as dipole signals in the residuals, or new physics such as non-gr polarization modes (Lee et al. 2008; Chamberlin & Siemens 2012) or massive gravitons (Lee et al. 2010) which would change the shape of the Hellings and Downs curve. We have, for the most part, also assumed that the intrinsic pulsar noise can be assumed to be white Gaussian noise with no discernible red noise. While previous work suggests that there will be red noise present in many Millisecond Pulsars (MSPs; Shannon & Cordes 2010), analyses of the present timing data (van Haasteren et al. 2011; D. Perrodin et al. 2013, in preparation; J. A. Ellis et al. 2013, in preparation) suggest that the data is white-noise-dominated and there is little to no evidence for red noise. However, further study of the model selection problem taking in to account the aforementioned effects is crucial to present detection efforts and will be the subject of a future paper.

4.2. The Detection Problem

We now turn to the question of detection. In a Bayesian analysis we would like to compute the odds that there is a GWB present in our data. Not surprisingly, the tool normally used to this end is the Odds ratio or the Bayes factor. Consider two models that we will label M_1 and M_2, then the Odds ratio is defined as

$$\mathcal{O} = \mathcal{B}(M_1, M_2|\mathbf{r})\frac{p(M_1)}{p(M_2)}, \tag{36}$$

where

$$\mathcal{B}(M_1, M_2|\mathbf{r}) = \frac{\int d\boldsymbol{\theta}_1\, p(\mathbf{r}|\boldsymbol{\theta}_1, M_1)p(\boldsymbol{\theta}_1)}{\int d\boldsymbol{\theta}_2\, p(\mathbf{r}|\boldsymbol{\theta}_2, M_2)p(\boldsymbol{\theta}_2)} \tag{37}$$

is the Bayes factor (i.e., the ratio of the marginalized likelihood functions over parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ corresponding to models M_1 and M_2, respectively), $\mathbf{r}$ is our data and $p(M_1)$ and $p(M_2)$ are the a priori probabilities on models M_1 and M_2, respectively. Note that the Bayes factor is the data-dependent part of the odds ratio where the a priori probabilities of the models is somewhat subjective, and as such, we will only consider Bayes factors when discussing detection in the this work.[4] For our purposes, we would like to compare at least three different models when weighing the odds of a stochastic GWB in our data:

1. $M_{\rm gw}$. A power law stochastic GWB with spatial correlations described by the Hellings and Downs coefficients $\zeta_{\alpha\beta}$, amplitude $A_{\rm gw}$ and power spectral index $\gamma_{\rm gw}$, individual power law red noise processes for each pulsar with amplitude A_α and power spectral index γ_α and white noise for each pulsar characterized by an EFAC parameter $\mathcal{F}_\alpha$ and EQUAD parameter $\mathcal{Q}_\alpha$.
2. $M_{\rm corr}$. A common red noise process among pulsars (as suggested in Shannon & Cordes 2010) with no spatial correlations and individual intrinsic red and white components as in model $M_{\rm gw}$.
3. $M_{\rm null}$. Only intrinsic red and white noise processes with no common red or white noise components among pulsars.

Comparing models $M_{\rm gw}$ and $M_{\rm null}$ will tell us whether or not there is evidence for any common red noise in our data but it will not necessarily tell us that this common noise is due to the stochastic GWB or some other common red noise source. Hence, a large Bayes factor $\mathcal{B}(M_{\rm gw}, M_{\rm null}|\mathbf{r})$ is necessary but not sufficient for detection. However, the comparison of models $M_{\rm gw}$ and $M_{\rm corr}$ can really give us information about the nature of the common red noise signal. As the two aforementioned models are identical except for the spatial correlations, a large Bayes factor $\mathcal{B}(M_{\rm gw}, M_{\rm corr}|\mathbf{r})$ will give us the odds that there is a common red noise process described spatial correlations $\zeta_{\alpha\beta}$. Since these *spatial* correlations are the signature of a stochastic GWB, the condition that this Bayes factor be large is both the necessary and sufficient condition for detection. In fact, this Bayes factor is closely related to signal-to-noise ratios in previous detection schemes (Jenet et al. 2005; Anholm et al. 2009; Yardley et al. 2011; S. Chamberlin et al. 2013, in preparation) that measure the significance of the cross correlations.

This first-order likelihood approximation has already been tested on the open and closed (J. A. Ellis et al. 2013, in preparation) IPTA Mock Data Challenge, where all challenges consisted of 130 data points per pulsar with 36 pulsars. For the closed data challenge, we have computed the Bayes factors mentioned in the previous section. In J. A. Ellis et al. (2013, in preparation) we have shown that we do indeed see very strong evidence for both a common red noise signal and a red noise signal with spatial correlations described by the Hellings and Downs coefficients. However, as we mentioned above, the Bayes factor $\mathcal{B}(M_{\rm gw}, M_{\rm null})$ is much larger than $\mathcal{B}(M_{\rm gw}, M_{\rm corr})$. For this reason, we expect that in analysis of real PTA data we will begin to see strong evidence for common red noise before we are able to see strong evidence for the expected cross correlations. In other words, as we gain more sensitivity, the first two terms in Equation (30) will dominate the likelihood function and the third term will only play a significant role as our sensitivity increases further. A full analysis of this feature along with projected sensitivity curves based on future pulsar timing campaigns and hardware upgrades will be explored in future work.

4.3. The Empirical Distribution Function

Here we will test the consistency and unbiasedness of our model through injections. Simply put, it is a type of hypothesis testing similar to the Kolmogorov–Smirnov test. In this test, the null-hypothesis that our analysis is internally consistent is accepted when for $x\%$ of realizations, the true injected parameter lies within the inner $x\%$ of the marginalized posterior distribution. A similar test was done recently in van Haasteren & Levin (2013) in *one* dimension through the use of the empirical distribution function (EDF). Here we will review this method and generalize it to two-dimensional marginalized posterior distributions. We define the inner high-probability region (HPR) of the two-dimensional marginalized posterior distribution as

$$\int_W p(\theta_1, \theta_2)d\theta_1 d\theta_2 = a$$
$$W = \{\theta_1, \theta_2 \in \mathbb{R} : p(\theta_1, \theta_2) > L_a\}, \tag{38}$$

where L_a is some value >0 unique to each a that corresponds to a curve of equal probability in the two-dimensional parameter space. In practice we lay down a grid in this two-dimensional parameter space and perform our search over the two parameters of interest (for the stochastic background we search over A and γ, the dimensionless strain amplitude and power spectral index of the GWB). We then define a set of points $\{A_i, \gamma_i\} \in \mathcal{S}_a$: $p(A_i, \gamma_i) > L_a$, that is to say we find all points in our grid that correspond to posterior values that lie inside our contour curve L_a. To determine if the injected values of $\{A_{\rm true}, \gamma_{\rm true}\}$ lie within the HPR we simply check to see if the injected values are consistent with the set $\mathcal{S}_a$. To do this we first define the complementary set to be $\bar{\mathcal{S}}_a$ such that points that are in this set are outside or the HPR. Now we define two χ^2 functions in the parameter space

$$\chi_a(A_i, \gamma_i)^2 = \left(\frac{A_i - A_{\rm true}}{A_{\rm true}}\right)^2 + \left(\frac{\gamma_i - \gamma_{\rm true}}{\gamma_{\rm true}}\right)^2 \tag{39}$$

$$\bar{\chi}_a(A_j, \gamma_j)^2 = \left(\frac{A_j - A_{\rm true}}{A_{\rm true}}\right)^2 + \left(\frac{\gamma_j - \gamma_{\rm true}}{\gamma_{\rm true}}\right)^2, \tag{40}$$

where $\{A_i, \gamma_i\}$ and $\{A_j, \gamma_j\}$ are elements of the sets $\mathcal{S}_a$ and $\bar{\mathcal{S}}_a$, respectively. Finally, we define the EDF as

$$F_k(a) = \frac{1}{k}\sum_{n=1}^{k}\Theta\big(\min\bar{\chi}_a^2 - \min\chi_a^2\big), \tag{41}$$

[4] It is possible to use astrophysical information such as the expected level of the stochastic background compared to our noise or the expectation number of single sources to construct the a priori probabilities. Here we will quantify our ignorance by considering equal a priori probabilities of all tested models.

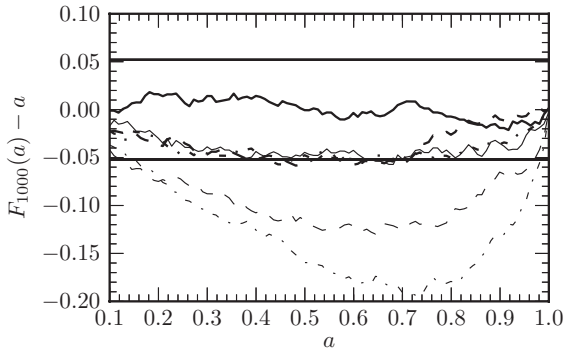

Figure 2. Empirical distribution function for six scenarios. The thick lines denote a 10 pulsar PTA and the thin lines denote a 15 pulsar PTA and the solid, dashed and dotted lines denote injected stochastic GWB amplitudes of 1×10^{-15}, 2×10^{-15}, and 3×10^{-15}, respectively. The solid lines at ±0.052 represent the value at which we should reject the null-hypothesis that our analysis method is consistent and unbiased.

where $\Theta(x)$ is the Heaviside function. The term inside the sum indicates an event when the injected values are "closer" (in the χ^2 sense) to one of the elements of $\mathcal{S}_a$ than to any of the elements of $\bar{\mathcal{S}}_a$, therefore we can say that the values $\{A_{\rm true}, \gamma_{\rm true}\}$ join the set $\mathcal{S}_a$ and lie within the HPR defined in Equation (38). Now that we have defined our EDF, the rest of the analysis mimics van Haasteren & Levin (2013).

For this analysis we simulated 1000 data sets for six different scenarios. In all cases we chose the white noise level to be 100 ns while we chose GWB amplitudes of 1×10^{-15}, 2×10^{-15}, and 3×10^{-15} for PTAs with both 10 and 15 pulsars with a 5 year baseline. Figure 2 shows the EDF for the six models outlined above.

The thick lines denote a 10 pulsar PTA and the thin lines denote a 15 pulsar PTA and the solid, dashed, and dotted lines denote injected stochastic GWB amplitudes of 1×10^{-15}, 2×10^{-15}, and 3×10^{-15}, respectively. The solid lines at ±0.052 represent the value at which we should reject the null-hypothesis that our analysis method is consistent and unbiased. First, we note that for both the 10 and 15 pulsar PTA, our analysis method is consistent for an injected amplitude of $A = 1\times10^{15}$. We obtain similar results in the 10 pulsar case for amplitudes of $A = 2\times10^{-15}$ and $A = 3\times10^{-15}$. Here we do see that our method is indeed slightly biased for these larger amplitudes but the degree of bias is almost negligible. However, for these same amplitudes in the 15 pulsar case there is a significant bias. Even though there is a bias present in these scenarios, the EDF does not give information about how this bias presents itself in the two-dimensional parameter space.

In Figure 3 we show the two-dimensional scatter plot of the maximum likelihood parameters from our Monte Carlo simulations. It is clear that the bias in our two-dimensional parameter space of interest is *practically* very small. In fact the means of the distributions for A and γ for the 10 pulsar case are $(1.6, 2.25, 3.14)\times10^{-15}$ and (4.17, 4.24, 4.23), respectively, and for the 15 pulsar case we obtain $(1.56, 2.29, 3.22)\times10^{-15}$ and (4.11, 4.12, 4.13), respectively. In the first row of Figure 3 we show the 10 pulsar case with increasing GWB amplitude and the second row we show the same for the 15 pulsar case. In the cases where there is a bias present, the likelihood function prefers slightly lower spectral indices and slightly larger amplitudes. However, from our experience with the Mock Data Challenge (MDC) this bias can also present itself by preferring a slightly higher spectral index and lower amplitude. It should be noted that even the smallest of the amplitudes tested here are near the upper range of the expected level of the stochastic GWB (Sesana 2012) and that the white noise rms of the pulsars is slightly unrealistic in our current PTA regime. In fact, we expect to have maybe five or six pulsars that time at or below the 100 ns level while we have many others that have much larger white noise rms. Thus we can conclude that even though our likelihood is somewhat biased at larger amplitudes (as is expected), for *realistic* astrophysically likely stochastic GWBs this method is effectively consistent and unbiased. In fact, in terms of setting upper limits on the stochastic GWB amplitude, this method is practically identical to using the full likelihood, while much more computationally efficient.

5. DISCUSSION AND CONCLUSIONS

Here will briefly discuss future prospects of conducting a simultaneous search for continuous GWs and the stochastic GWB. We will also compare our work to other recent efforts to speed up PTA GW data analysis and discuss the importance of our first-order likelihood method.

5.1. Simultaneous Detection of Continuous GWs and a Stochastic GWB

One very important feature of the first-order likelihood method is that it can also be applied to searches for continuous GWs. This will allow us to simultaneously search for a correlated stochastic background and resolve individual sources that are bright enough to stand out above such a background. In standard continuous GW searches using PTAs (Babak & Sesana 2012; Ellis et al. 2012b; Petiteau et al. 2013) the assumption is made that any detectable single source will be bright enough such that the noise (e.g., stochastic GWB) can be approximated as a Gaussian process that is uncorrelated among pulsars. However, recent work (Ravi et al. 2012) has shown that we are likely to see a few single sources per frequency bin that will stand out from the typical isotropic stochastic background, thus in order to resolve the weakest of these it is crucial to simultaneously search for a correlated stochastic background as well as the continuous source. We can then write down a combined likelihood function assuming a deterministic source of functional form $\mathbf{s}(\boldsymbol{\lambda})$

$$p(\mathbf{r}|\boldsymbol{\theta},\boldsymbol{\lambda}) = \frac{1}{\sqrt{\det 2\pi\,\boldsymbol{\Sigma}}}\exp\left(-\frac{1}{2}(\mathbf{r}-\mathbf{s})^T\boldsymbol{\Sigma}^{-1}(\mathbf{r}-\mathbf{s})\right), \quad (42)$$

where our noise (including the stochastic background) parameters are $\boldsymbol{\theta}$ and our single source parameters are $\boldsymbol{\lambda}$. Using our first-order likelihood approach, we can approximate Equation (42) as

$$\begin{aligned}\ln p(\mathbf{r}|\boldsymbol{\theta},\boldsymbol{\lambda}) = &\approx -\frac{1}{2}\left[\mathrm{Tr}\ln\mathbf{P} + (\mathbf{r}-\mathbf{s})^T\mathbf{P}^{-1}(\mathbf{r}-\mathbf{s})\right.\\ &\left.-(\mathbf{r}-\mathbf{s})^T\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1}(\mathbf{r}-\mathbf{s})\right]\\ &= -\frac{1}{2}\sum_{\alpha=1}^{M}\left[\mathrm{Tr}\ln P_\alpha + (r_\alpha - s_\alpha)^T P_\alpha^{-1}(r_\alpha - s_\alpha)\right.\\ &\left.-\sum_{\beta\neq\alpha}^{M}(r_\alpha - s_\alpha)^T P_\alpha^{-1} S_{\alpha\beta} P_\beta^{-1}(r_\beta - s_\beta)\right]. \quad (43)\end{aligned}$$

As in the stochastic background case, this again will speed up computations because we only have to invert the *individual* auto-covariance matrices as opposed to the *full* data covariance matrix. Although there have been proposed methods to speed

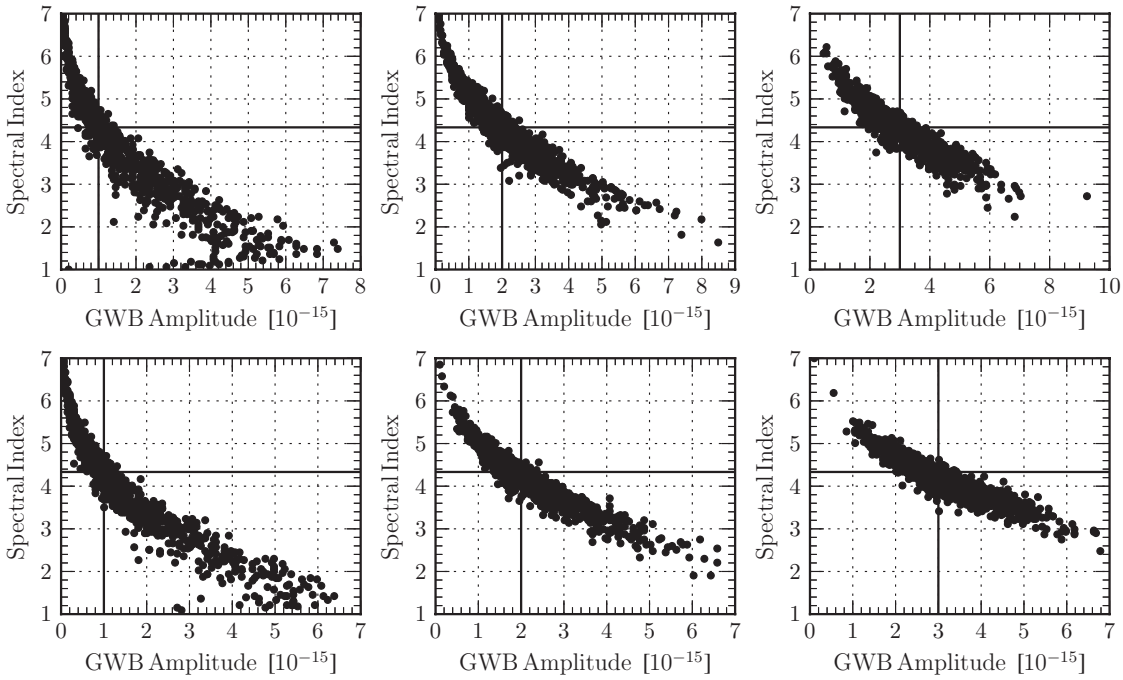

Figure 3. Here we show the scatter of the maximum likelihood values of the GWB amplitude and spectral index from the Monte Carlo simulations. From left to right the injected amplitudes are 1×10^{-15}, 2×10^{-15}, and 3×10^{-15} with spectral index 13/3 for a 10 pulsar PTA (top row) and 15 pulsar PTA (bottom row). We can see that nearly all of these distributions display minimal bias.

up the computation of the stochastic likelihood function of Equation (15) (van Haasteren 2013), this is not applicable to continuous sources because it relies on essentially applying a low pass filter to the data. However, since we expect continuous sources across the entire frequency band (with higher frequency sources possibly standing out above the background) we must keep all frequency information. Therefore our first-order likelihood approximation is a viable option when looking to significantly speed up computation time while losing minimal information about potential GW signals.

As always, to claim a detection we must do some sort of model comparison, be it a Neyman-Pearson test for Frequentist statistics or an odds ratio or Bayes factor for Bayesian statistics. For example if we want to assess the likelihood of that a continuous GW is in our data we want to compute the following Bayes factor

$$\mathcal{B} = \frac{\mathcal{Z}_{\rm CW}}{\mathcal{Z}_{\rm noise}} = \frac{\int\int d\boldsymbol{\lambda} d\boldsymbol{\theta}\, p(\mathbf{r}|\boldsymbol{\theta},\boldsymbol{\lambda}) p(\boldsymbol{\lambda}) p(\boldsymbol{\theta})}{\int d\boldsymbol{\theta}\, p(\mathbf{r}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}, \qquad (44)$$

where $\mathcal{Z}_{\rm CW}$ and $\mathcal{Z}_{\rm noise}$ are the evidence for the gravitational Continuous Wave (CW) and noise models, respectively. However, note that $\boldsymbol{\theta}$ depends on our stochastic GWB parameters as we treat all stochastic processes as "noise" in this analysis. If we do not include the GWB parameters in the model then we could mistake a low frequency GWB for a single continuous source, thus including the GWB stochastic background in both models is crucial to detection and eventually characterization of a single GW source. We should also mention that the biases mentioned in Section 4.3 are not as important if we simply wish to let the noise parameters vary along with the single source parameters since these noise parameters will be marginalized over in the end. An exploration of these combined searches will be the subject of a future paper.

5.2. Comparison with Other Work

Recently there have been three studies devoted to making the analysis of PTA data more computationally efficient. First, van Haasteren (2013, hereafter vH13) have developed a method called Acceleration By Compression (ABC) to speed up this analysis. The main point of that work is to write the data in a compressed basis, keeping the minimum number of basis vectors to maximize the ability to characterize a correlated red signal. vH13 also makes use of an interpolation scheme to compute the covariance matrix which further improves the efficiency of the algorithm at the cost of large memory usage. The aforementioned method has proved to be very efficient and accurate in setting upper limits on the stochastic GWB and characterizing injected signals, however, since it relies on a reduced basis that essentially "throws away" high frequency information it is impossible to obtain a reliable Bayes factor when comparing models that allow for varying white noise components. Since our first-order likelihood function makes use of all the information in the data we can indeed compute reliable Bayes factors and make confident statements about detection. We note, however, that the first-order likelihood of this work and the ABC method of vH13 are complementary. The two methods can in principle be combined for even greater efficiency.

Most recently there have been two analyses of the IPTA MDC that aim to make the PTA data analysis more efficient. First, Lentati et al. (2012) have developed a novel model-independent method for the estimation of the spectral properties of an isotropic stochastic GWB. It makes use of a frequency domain approach and is extremely efficient and results in computational speedups of two to three orders of magnitude over the full likelihood implementation. It has also been extensively tested on the MDC data sets and has proved to be very accurate in characterizing the stochastic GWB. Our first-order likelihood method is indeed complementary to the aforementioned work as

it provides a way to efficiently evaluate the likelihood function in a full time domain analysis which will be vital for cross-checks of real-life detection candidates.

Finally, Taylor et al. (2013) have implemented the full VHML likelihood function and have made it more efficient through the use of optimized linear algebra libraries with multithreading and parallelization resulting in significant speedups in the likelihood evaluation. However, all of these methods could just as well be applied to the first-order likelihood which would still be more efficient than the full likelihood by a factor proportional to the number of pulsars in the array.

This work and recent work have shown that there has indeed been significant progress on making the likelihood evaluation more efficient for PTAs. All of these methods are complementary and will provide important cross checks for future stochastic GWB detection candidates.

5.3. Summary

In this paper we have introduced a novel way to speed up the computation of the likelihood function for PTAs when searching for a stochastic GWB. This was accomplished by expanding the likelihood function to first order in the Hellings and Downs correlation coefficients expected for a stochastic GWB leading to a computational speedup on the order of the square of the number of pulsars in the PTA. For typical PTAs this results in a speed-up of a few hundred to about a thousand. We have briefly discussed the implementation of this technique on the first IPTA Mock Data Challenge and showed that this algorithm performs well in extracting the injected GWB parameters and making a significant detection through various Bayes factors. Though this is indeed an approximation to the full likelihood function we have shown through extensive simulations that the bias introduced in the estimation of GWB parameters is minimal and negligible in many cases. This was accomplished through an analytical computation of the expectation value of the maximum likelihood, direct comparisons of the full and first-order likelihood functions on simulated data sets and through a statistical Monte Carlo approach based on the EDF. Although this work has focused solely on the detection and characterization of a stochastic GWB, this likelihood function can also be used to estimate the intrinsic red and white noise parameters of individual pulsars simultaneously with the GWB parameters.

We thank the members of the NANOGrav detection working group for their comments and support, especially Paul Demorest and Joe Romano. We also thank Jolien Creighton for useful conversations. This work was partially funded by the Wisconsin Space Grant Consortium and the NSF through CAREER award No. 0955929, PIRE award No. 0968126, and award No. 0970074.

APPENDIX

RELATIONSHIP TO VHML LIKELIHOOD

Making use of Equation (6), the likelihood function for the noise can be written as

$$p(\mathbf{n}|\boldsymbol{\theta}) = p(\mathbf{r}|\boldsymbol{\theta}, \delta\boldsymbol{\xi}_{\rm best}) = \frac{1}{\sqrt{\det(2\pi\,\boldsymbol{\Sigma}_n)}} \times \exp\left(-\frac{1}{2}(\mathbf{r} - \mathbf{M}\delta\boldsymbol{\xi}_{\rm best})^T\,\boldsymbol{\Sigma}_n^{-1}(\mathbf{r} - \mathbf{M}\delta\boldsymbol{\xi}_{\rm best})\right). \quad \text{(A1)}$$

This can be thought of as a *conditional* pdf, where the values of $\delta\boldsymbol{\xi}_{\rm best}$ are fixed. In van Haasteren & Levin (2013) it was shown that the marginalized likelihood can be written as

$$p(\mathbf{r}|\boldsymbol{\theta}) = \int d\delta\boldsymbol{\xi}\; p(\mathbf{r}|\boldsymbol{\theta}, \delta\boldsymbol{\xi}) = \frac{\exp\left[-\frac{1}{2}\mathbf{r}^T\mathbf{G}^T(\mathbf{G}^T\,\boldsymbol{\Sigma}_n\mathbf{G})^{-1}\mathbf{G}^T\mathbf{r}\right]}{\sqrt{\det 2\pi\,\mathbf{G}^T\,\boldsymbol{\Sigma}_n\mathbf{G}}}, \quad \text{(A2)}$$

where $\mathbf{G}$ is the matrix constructed from the final $(N - N_{\rm fit})$ columns of the matrix $\mathbf{U}$ in the SVD of the design matrix, $\mathbf{M} = \mathbf{USV}^T$.

We will now explore the G matrix and the R matrix obtained from the marginalized and conditional pdfs, respectively. As mentioned above, R can be thought of as an oblique projection operator that projects the pre-fit residuals into the post-fit residual space, whereas G^T can be thought of a projection operator that projects our data onto the null space of M, that is, it projects the data into a subspace orthogonal to the timing model fit. Since R is not generally symmetric and therefore is an oblique projection operator, it does not have such a simple mathematical interpretation. However, we can recast our problem in terms of "weighted" residuals then we have the following transformations: $r \rightarrow Wr$, $M \rightarrow WM$, and $R \rightarrow W^{-1}RW$, where W is the weighting matrix defined above. In this case minimizing the χ^2 becomes an unweighted least squares problem and we obtain the exact same estimates of $\delta\boldsymbol{\xi}_{\rm best}$ and likelihood function as before. In this case R is symmetric and can be thought of as an orthogonal projection operator that projects our weighted data onto the null space of the weighted timing model (WM). However, in order to compute the likelihood we still have to invert the covariance matrix $\Sigma_r = R\Sigma_n R^T$ which is singular. To do this we rely on the pseudo-inverse. The pseudo-inverse of Σ_r is easiest defined in terms of its eigen-decomposition $\Sigma_r = EDE^T$, with E the matrix of eigenvectors of Σ_r, and D the diagonal matrix with $D_{ii} = \lambda_i$ the eigenvalues of Σ_r. It so happens that for a symmetric positive semi-definite matrices like these, the eigen-decomposition is also the SVD. The pseudo-inverse of Σ_r is then

$$\overline{\Sigma_r^{-1}} = E\overline{D^{-1}}D^T, \quad \text{(A3)}$$

where the overbar indicates that we are taking a pseudo-inverse and $\overline{D^{-1}}_{ii} = 1/\lambda_i$ for $\lambda > 0$ and $\overline{D^{-1}}_{ii} = 0$ otherwise. Note that when all the error bars are the same (i.e., $W = \sigma^{-1}\mathbb{I}$ with σ constant), the matrix $G^T\Sigma_n G$ has the same eigenvalues as the non-singular part of $R\Sigma_n R^T$ and we have

$$\overline{(R\Sigma_n R^T)^{-1}} = G(G^T\Sigma_n G)^{-1}G^T. \quad \text{(A4)}$$

Thus we have obtained a very interesting result that in the case of uniform uncertainties, the conditional pdf making use of a pseudo-inverse is equivalent to the marginalized pdf making use of the projection matrix G^T. However, in general this is not true and the two methods are indeed different. Although, in many cases the uncertainties are similar on a majority of the TOAs, thus the two methods will not differ much in practice.

REFERENCES

Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., & Siemens, X. 2009, PhRvD, 79, 084030
Babak, S., & Sesana, A. 2012, PhRvD, 85, 044034

Chamberlin, S. J., & Siemens, X. 2012, PhRvD, 85, 082001
Coles, W., Hobbs, G., Champion, D. J., Manchester, R. N., & Verbiest, J. P. W. 2011, MNRAS, 418, 561
Corbin, V., & Cornish, N. J. 2010, arXiv:1008.1782
Cordes, J., & Jenet, F. 2012, ApJ, 752, 54
Cordes, J. M., & Shannon, R. M. 2010, arXiv:1010.3785
Cordes, J., & Shannon, R. 2012, ApJ, 750, 89
Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94
Detweiler, S. 1979, ApJ, 234, 1100
Ellis, J. A., Jenet, F. A., & McLaughlin, M. A. 2012a, ApJ, 753, 96
Ellis, J. A., Siemens, X., & Creighton, J. D. E. 2012b, ApJ, 756, 175
Enoki, M., Inoue, K. T., Nagashima, M., & Sugiyama, N. 2004, ApJ, 615, 19
Finn, L. S., & Lommen, A. N. 2010, ApJ, 718, 1400
Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39
Hobbs, G., Archibald, A., Arzoumanian, Z., et al. 2010, CQGra, 27, 084013
Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655
Jaffe, A. H., & Backer, D. C. 2003, ApJ, 583, 616
Jenet, F. A., Hobbs, G. B., Lee, K. J., & Manchester, R. N. 2005, ApJL, 625, L123
Jenet, F. A., Hobbs, G. B., van Straten, W., et al. 2006, ApJ, 653, 1571
Jenet, F. A., Lommen, A., Larson, S. L., & Wen, L. 2004, ApJ, 606, 799
Lee, K. J., Jenet, F. A., & Price, R. H. 2008, ApJ, 685, 1304
Lee, K., Jenet, F. A., Price, R. H., Wex, N., & Kramer, M. 2010, ApJ, 722, 1589
Lee, K. J., Wex, N., Kramer, M., et al. 2011, MNRAS, 414, 3251
Lentati, L., Alexander, P., Hobson, M. P., Taylor, S., & Balan, S. T. 2012, arXiv:1210.3578
Lommen, A. N. 2002, in Neutron Stars, Pulsars, and Supernova Remnants, ed. W. Becker, H. Lesch, & J. Trümper (Garching bei München, Germany: Max-Plank-Institut für extraterrestrische Physik), 114
Lommen, A. N., & Backer, D. C. 2001, ApJ, 562, 297
McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2012, arXiv:1211.5377
Mingarelli, C., Grover, K., Sidery, T., Smith, R., & Vecchio, A. 2012, PhRvL, 109, 081104
Petiteau, A., Babak, S., Sesana, A., & de Araujo, M. 2013, PhRvD, 87, 064036
Ravi, V., Wyithe, J. S. B., Hobbs, G., et al. 2012, ApJ, 761, 84
Roedig, C., & Sesana, A. 2012, JPhCS, 363, 012035
Sesana, A. 2012, arXiv:1211.5375
Sesana, A., & Vecchio, A. 2010, PhRvD, 81, 104008
Sesana, A., Vecchio, A., & Colacino, C. N. 2008, MNRAS, 390, 192
Sesana, A., Vecchio, A., & Volonteri, M. 2009, MNRAS, 394, 2255
Shannon, R. M., & Cordes, J. M. 2010, ApJ, 725, 1607
Stinebring, D. R., Ryba, M. F., Taylor, J. H., & Romani, R. W. 1990, PhRvL, 65, 285
Taylor, J., Wolszczan, A., Damour, T., & Weisberg, J. 1992, Natur, 355, 132
Taylor, S. R., Gair, J. R., & Lentati, L. 2013, PhRvD, 87, 044035
van Haasteren, R. 2013, MNRAS, 429, 55
van Haasteren, R., & Levin, Y. 2010, MNRAS, 401, 2372
van Haasteren, R., & Levin, Y. 2013, MNRAS, 428, 1147
van Haasteren, R., Levin, Y., Janssen, G. H., et al. 2011, MNRAS, 414, 3117
van Haasteren, R., Levin, Y., McDonald, P., & Lu, T. 2009, MNRAS, 395, 1005
Volonteri, M., Haardt, F., & Madau, P. 2003, ApJ, 582, 559
Waldman, S. 2011, arXiv:1103.2728
Wyithe, J. S. B., & Loeb, A. 2003, ApJ, 590, 691
Yardley, D. R. B., Coles, W. A., Hobbs, G. B., et al. 2011, MNRAS, 414, 1777
Yardley, D. R. B., Hobbs, G. B., Jenet, F. A., et al. 2010, MNRAS, 407, 669